KW

RESTRICTED CIRCULATION

This paper is not an official document; it is for information only and must not be quoted in any printed publication.

Agricultural Memorandum No 983

Creating the 'Complete' Agromet Dataset

by M Field

1. The need for a 'complete' or 'bogus' Agromet Dataset

The climatic limitations of the revised Agricultural Land Classification Scheme are assessed in terms of Accumulated Temperature and Potential Soil Moisture Deficit (see Agricultural Memorandum No 984). To calculate these climatic parameters daily records of the following elements are required for the 20-year period 1961-80 for a number of stations in England and Wales -

(a) rainfall amount

(b) windspeed

(c) sunshine duration

(d) vapour pressure

(e) maximum and minimum temperature

The Agromet Dataset is suitable for such an application. It consists of daily data, climatologically ordered, for 163 stations in the UK. However not all of the stations cover the period 1961-80. For many stations the data start in 1959 but there is considerable variation, for example Hadlow College (1966-82) and Thorney Island (1957-76). Moreover even for the stations which cover the period 1961-80 very few have complete records of all 6 elements. For some stations one or more of the elements are not recorded at all; this applies in particular to sunshine and windspeed. There are also gaps of various lengths in the records due to such factors as missing observations and temporary closure of a station.

One could arrange for the programs which calculate Accumulated Temperature and Potential SMD to substitute for missing data using the records from a nearby station(s). However there are dangers in this. For example suppose there are some missing sunshine data from an inland station which is at an altitude of 200m. The only nearby station may be on the coast which will be quite unsuitable as a straight substitute. The only satisfactory solution to the problem of missing data is to use an objective method which takes account of the spatial variability of the meteorological element. Such a technique has been devised by Tabony, see (4) and (5). Once this technique has been applied and the dataset has been 'completed' it will find other applications for example in the Long Term Irrigation Planning Program (see Ag. Mem. 947).

2. The method of 'completing' the dataset

The first task was to select from the 163 stations those that were open during the period 1961-80. Having done this and plotted their positions on a map it was apparent that there were a number of 'holes' in the network. Some of the holes were simply due to the paucity of observing stations in certain parts of the country such as in Cumbria and the central part of South Wales - nothing could be done about this. However a few holes could be filled in by accepting stations that were not open for the full 20-year period. These were Leeming (1966-80), Coltishall (1963-80) and Yeovilton (1965-80).

The missing data could be divided into 2 classes,

(i) those data that were missing for the entire period (61-80), eg a missing wind record because there was no anemometer, and

(ii) those data that were missing intermittently due to missing observations, etc.

Class (i) data were completed first, followed by Class (ii). For a discussion of the windspeed problem see later.

2.1 Class (i) some station had no record of sunshine and/or vapour pressure (all stations had a record, albeit a broken one, of max and min temperature and rainfall).

A nearby station was selected which had a complete record of the element required, care being taken that the station was well-exposed and, as far as one could tell, had data of good quality. Each of the 7305 missing daily values over the 20 year period was estimated by multiplying the value at the 'replacement' station by a factor. This factor was the ratio of the long-term mean annual value of the element at the 'incomplete' and 'replacement' stations. The long-term means were found from maps published in various memoranda by the Climatological Services Branch (see References and Appendix I). The positions of the two sites were plotted on the maps from which the long-term means could be estimated.

2.2 Class (ii)

Rainfall - the nearest raingauge to the station was found and its data were inserted into the missing records multiplying by the ratio of the 1941-70 Average Annual Rainfalls at the two sites. The objective method described below could not be used very easily because of the arrangement of the rainfall datasets.

Sunshine, vapour pressure, max and min temperature - a technique was needed in which those data that were present at a station were compared with the corresponding data at a number of stations around it. The 'satellite' station giving the best correlation would then be used to estimate objectively for missing values. A program to do this already existed, written by Tabony of the Climatological Services Branch, but required some modification. Listings of the original and the modified subroutines with some additional documentation are held in the Agricultural Branch.

The program assumed that the data belonged to one population and so it operated on one month at a time only. Thus for a particular element all the missing data for January were completed first then all of February's, and so on in a loop. It would accept up to 200 stations at a time but since the amount of work the program did was roughly proportional to the square of the number of stations the cost in computer units of using them all was prohibitive. Moreover there was little merit in looking for good correlations between stations at opposite ends of the country. Thus England and Wales were divided into clusters of stations and the program was run separately on each cluster. Arbitrarily there were 15 stations per cluster. In each run of the program one element was considered and the data for all 15 stations were made complete for the whole period.

The program operated in 2 stages,

Stage 1 a matrix of Fisher's 'Z statistic' was produced so that each of the 15 stations could be ranked with its 14 satellites.

Stage 2 the data from each station was considered in turn with the corresponding data from its 'best satellite' to produce a regression equation which could be used to estimate for missing values. If there were not at least 150 days of overlapping data for the month in question the 'next-best' satellite was used instead. Fewer than 150 data points was not considered to be sufficient since there were always other stations available which gave greater overlap. The 'best' or 'next best' satellite was readily found from the ranking of Stage 1. Further details are given in (4).

One of the modifications that was made involved the Regression technique. Consider the scatter diagram for sunshine,

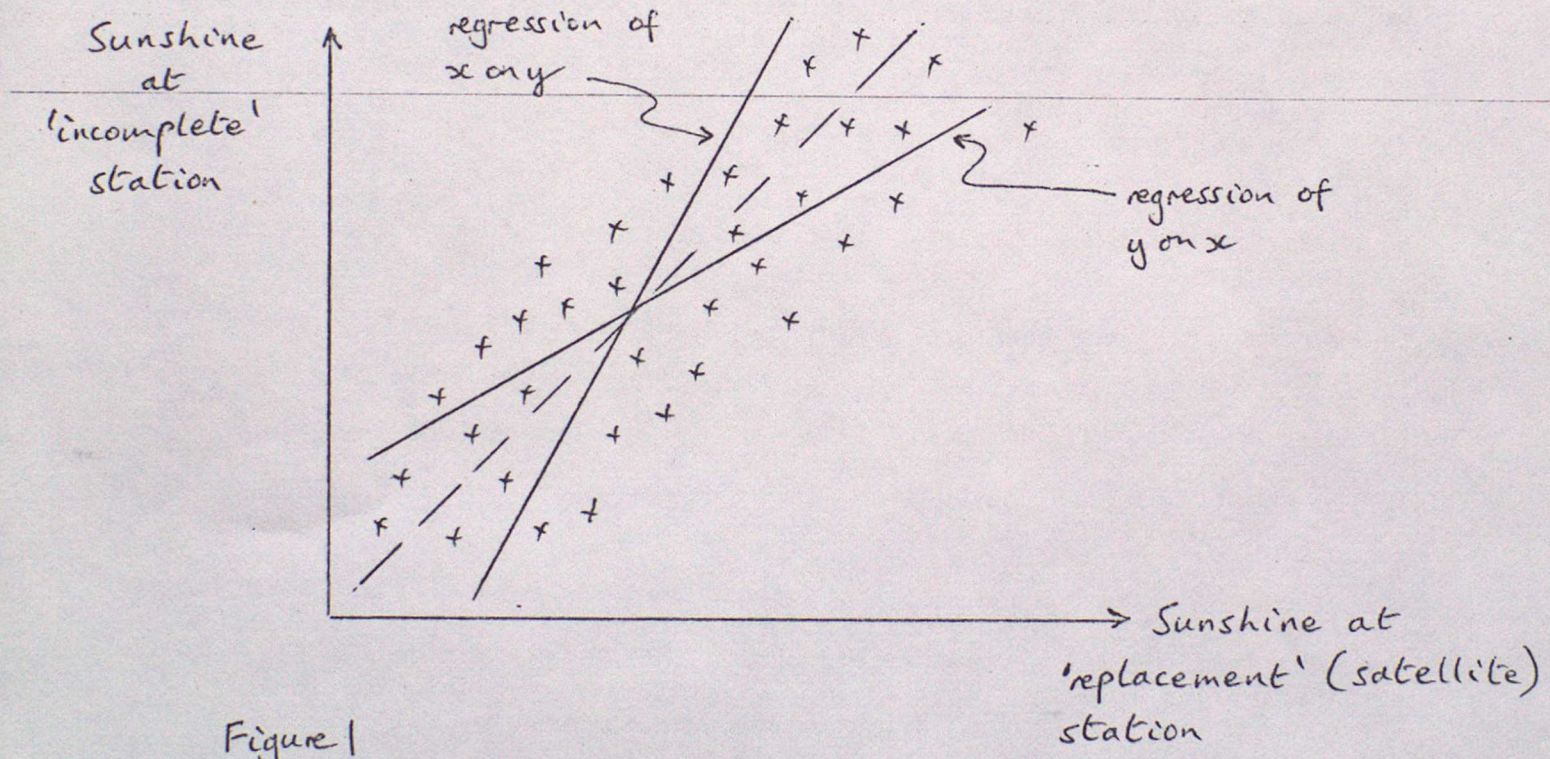


Figure 1

The conventional approach is to do a regression of y on x so that the quantity $\sum_{i=1,N} (y_o - y_e)^2$ is a minimum, where y_o is the observed value and y_e is the estimated value. However this will give a slope which is < 1 as shown. The regression of x on y gives a slope which is > 1. In fact if the points were like currants in a pudding we would get the following,

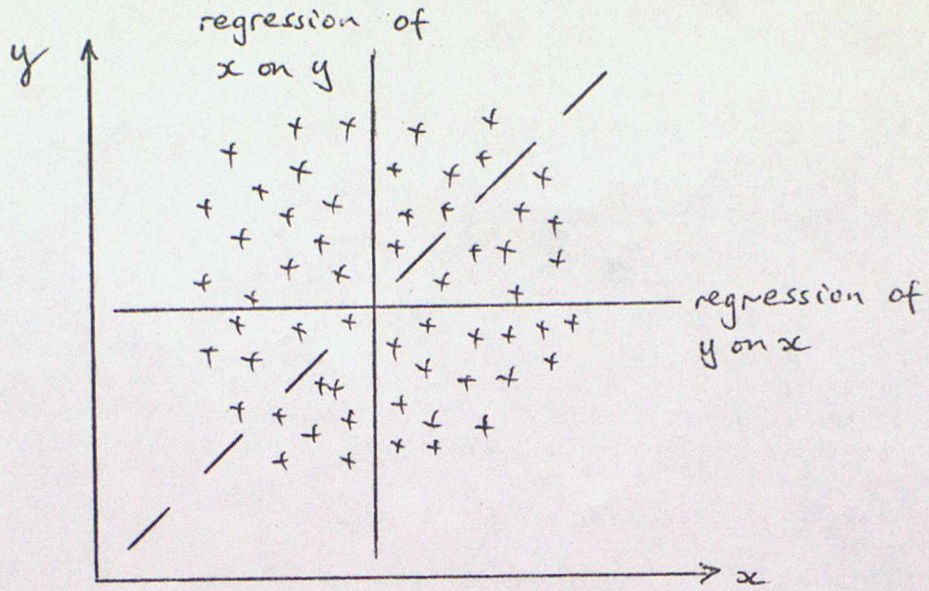


Figure 2

By having an intercept on the y-axis (see Fig 1) we always estimate small, positive values when the data at the satellite station are zero. This can lead to results which look a little odd. For example if there is a week with no sunshine in an area the satellite station will have 7 values of 0.0 hours but the estimated values may come out at 7 x 0.1 hours. It would seem more reasonable to fit the regression between that of x on y and y on x, ie 'down the middle' as indicated by the dashed lines in Figures 1 and 2. This way the slope would be approximately 1.0 (sometimes more, sometimes less) and not systematically less than 1.0.

To regress y on x the equation is $y - \bar{y} = m(x - \bar{x})$ where $m = \sigma_y r / \sigma_x$.
r = correlation coefficient = $cov(x,y)/\sigma_x\sigma_y$

(to regress x on y σ_x and σ_y are interchanged in the above equations)

To fit the dashed line the equation is $y - \bar{y} = m'(x - \bar{x})$ where $m' = \sigma_y / \sigma_x$

To get a rough idea of how well the technique worked a station was selected which had complete records of max and min temperature, sunshine and vapour pressure. On several arbitrary days during the 20-year period for the 2 months of February and July missing data (-32768) were inserted into the record. The program was run and the estimated values were compared with the original from which the mean error of the daily estimates would be deduced (see Table 1). The long-term monthly means (41-70) are also supplied to help interpret the results.

Element	Feb	July
Max temp (°C)	0.5 (7.0)	0.4 (21.5)
Min temp (°C)	0.8 (1.0)	1.0 (12.0)
Sunshine (hour)	0.7 (2.0)	0.15 (5.5)
Vapour pressure (mb)	0.3 (7.0)	0.8 (13.0)

Table 1 Mean error of daily estimates for 4 elements with long-term monthly means in brackets. (Data from Elmdon).

Errors in min temperature were larger than for max temperature because the minimum exhibits greater spatial variability. This is due to the effects of local topography and soil type on radiation nights. There was little difference in the error in max or min between February and July.

Errors in sunshine duration were greater in winter than summer due to factors such as fog clearance which is often highly localized. In the summer sunshine is spatially much more coherent than in winter, hence the lower error.

Vapour Pressure is known to be quite coherent spatially; the error appeared to be about 5% of the mean value in both February and July.

2.3 Windspeed - The situation regarding missing wind data was more complicated than for the other elements because wind measurements are made at various heights above ground level, 2 m and 10 m being the most common. Before attempting to 'complete' the records the data had first to be 'standardized', i.e. all the wind data expressed at a common height such as 10 m. The following problems were encountered in trying to do this.

(a) Anemometers have a starting speed which is of the order of a few knots, according to the instrument type and condition. Thus an instrument exposed at 2 m will spend more of its time stationary than if it were at 10 m where the wind is stronger. The quantitative effect of this is hard to assess but the important point is that the 2 m instrument will give a greater systematic underestimate of windspeed than the 10 m one. Theoretical correction factors such as Hellmann's (9) do not account for instrumental factors such as this.

(b) The units in which the windspeed was measured were often not clear. This applied particularly to 2 m cup-counter run-of-wind records where there is confusion over whether the units are miles per day or km per day. It has not been possible to resolve this problem.

(c) The exposure of 2 m cup-counters is often not as good as 10 m anemometers. The latter are frequently found on well-exposed airfields whereas cup-counters are often sheltered by hedges, trees, etc. The cup-counter at Plumpton for example is adjacent to a wind-break.

(d) At some stations there is an anemograph at 10 m or thereabouts as well as a cup-counter (eg Preston Wynne). It is often not clear from which of the instruments the record on the dataset has been compiled.

For the above reasons it was considered an impossible task to convert the wind data to a common height above ground level. In fact only the wind data from the main meteorological observing stations with anemometers at 10 m were considered to be of acceptable quality. Data from all other stations were thus ignored.

To provide a complete 10 m wind record for all 94 sites was a twofold task,

(i) To 'complete' the 10 m wind data for the main meteorological stations which were to act as 'bench marks' (there were 16 of these),

(ii) To use these stations to infer a wind record for each of the 78 remaining sites.

For steps (i) and (ii) the ratio of the long-term means at the 2 stations was needed - this factor was deduced from Climatological Memorandum 72. In step (i) a station with a complete record such as Elmdon was used to complete the data at the 'incomplete' station by applying the factor only on those days when there were missing data. In step (ii) the entire record was transferred from the 'replacement' station to the 'incomplete' station multiplying by the factor on each of the 7305 days over the 20-year period.

3. The 'complete' Agromet Dataset There are 94 stations on the 'complete' dataset, as follows -

Aberporth (Dyfed)
Askham Bryan (N. Yorkshire)
Aylesbury (Buckinghamshire)
Boscombe Down (Wiltshire)
Bradford (W Yorkshire)
Bramham (W Yorkshire)
Bude (Cornwall)
Buxton (Derbyshire)
Cambridge BG (Cambridgeshire)
Cambridge NIAB (Cambridgeshire)
Cardington (Bedfordshire)
Carlisle (Cumbria)
Cawood (N Yorkshire)
Cheltenham (Gloucestershire)
Cockle Park (Northumberland)
Coltishall* (Norfolk)
Cranwell (Lincolnshire)
Crumbland (Gwent)
Cwmystwyth (Dyfed)
Dale Fort (Dyfed)
Durham (Durham)
Eastbourne (E Sussex)
East Malling (Kent)
Elmdon (W Midlands)
Everton (Hampshire)
Faversham (Kent)
Fernhurst (W Sussex)
Filton (Avon)
Finningley (S Yorkshire)
Gatwick (W Sussex)
Gloucester (Gloucestershire)
Goggerdan (Dyfed)
Gorleston (Norfolk)
Goudhurst (Kent)
Hastings (E Sussex)
Hawkridge (Somerset)
Haydon Bridge (Northumberland)
Heathrow (Greater London)
Helmshore (Lancashire)
High Mowthorpe (N Yorkshire)
Hoddesdon (Hertfordshire)
Hull (Humberside)
Hurley (Berkshire)
Hurn (Dorset)
Keele (Staffordshire)
Kew (Greater London)
Kirkham (Lancashire)
Leeming* (N Yorkshire)
Loggerheads (Clwyd)
Long Ashton (Avon)
Long Sutton (Hampshire)
Lowestoft (Suffolk)
Lyneham (Wiltshire)
Lyonshall (Hereford & Worcester)
Manston (Kent)
Martyr Worthy (Hampshire)
Mount Batten (Devon)
Newport (Salop)
Newton Rigg (Cumbria)
Oxford (Oxfordshire)
Pen-y-Ffridd (Gwynedd)
Pershore (Hereford & Worcester)
Porton (Wiltshire)
Preston Wynne (Hereford & Worcester)
Rhoose (S Glamorgan)
Ringway (Greater Manchester)
Rosewarne (Cornwall)
Rothamsted (Hertfordshire)
Rustington (W Sussex)
Santon Downham (Norfolk)
Shawbury (Salop)
Shoeburyness (Essex)
Silpho Moor (N Yorkshire)
Silsoe (Bedfordshire)
Skegness (Lincolnshire)
Slaidburn (Lancashire)
Stansted (Essex)
Starcross (Devon)
Stratford-upon-Avon (Warwickshire)
Sutton Bonington (Nottinghamshire)
Swansea (Dorset)
Terrington St Celement (Norfolk)
Trawscoed (Dyfed)
Valley (Gwynedd)
Warsop (Nottinghamshire)
Wattisham (Suffolk)
Wellesbourne (Warwickshire)
Wisley (Surrey)
Wittering (Cambridgeshire)
Woburn (Bedfordshire)
Writtle (Essex)
Wye (Kent)
Wyton (Cambridgeshire)
Yeovilton* (Somerset)

* short-period station

Acknowledgements

The helpful suggestions and advice given by R Tabony are gratefully acknowledged. My thanks too to L Harvey and A Woodcock without whose hard work the dataset would not have been finished on time.

Any comments or suggestions on this paper should be referred to the author.

Meteorological Office, Met O 8a
London Road
Bracknell
Berkshire

July 1983

References

1. Climatological Memorandum No. 72, Maps of Average Duration of Bright Sunshine over the United Kingdom, 1941-70. Published by Climatological Services Branch, Meteorological Office, Bracknell, Berks. Dec 1974.

2. Climatological Memorandum No. 75, Maps of mean Vapour Pressure and of Frequencies of High Relative Humidity over the United Kingdom. September 1975.

3. Climatological Memorandum No. 79, Maps of Hourly mean windspeed over the United Kingdom, 1965-73, by Dr P G F Caton, August 1976.

4. Subroutine MISSIN, by R Tabony (a computer program write-up published within the Climatological Services Branch of the Meteorological Office, Bracknell).

5. The Estimation of Missing Climatological Data. Met O 3 Technical Note 8, R Tabony.

6. Field M, Agricultural Memorandum No 947, 'Long Term Irrigation Planning - the Computer Program Approach', Oct 1982.

7. Barrie I A, Agricultural Memorandum No 962, 'A Comparison of windspeeds recorded at 10 m and 2 m', October 1982.

8. Field M, Agricultural Memorandum No 984, 'Agricultural Land Classification - some meteorological aspects of the revised scheme', July 1983.

9. Meteorological Office Observer's Handbook, 3rd edition 1969, P75.

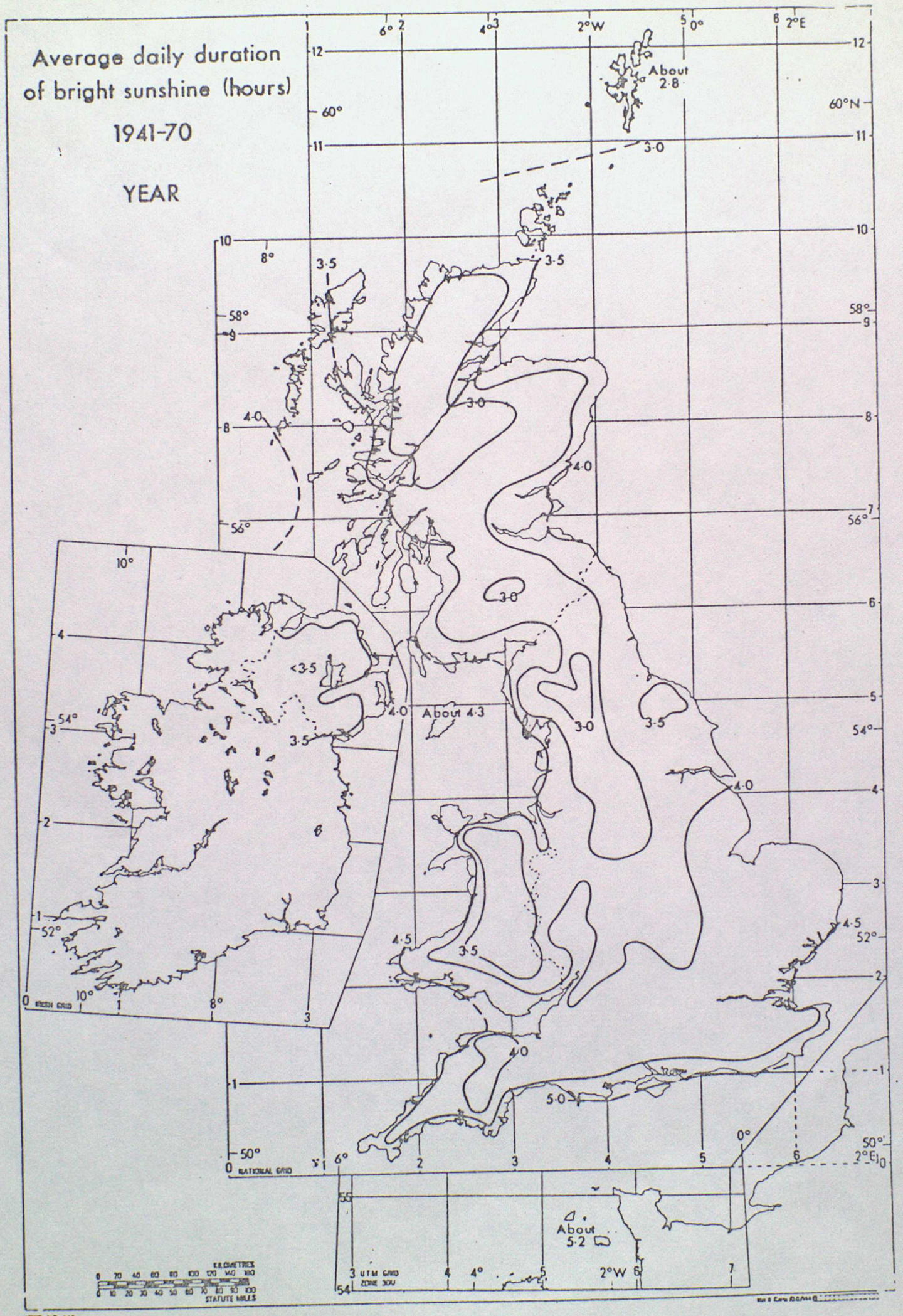
NB Copies of Climatological Memoranda and Technical Note 8 are available from:

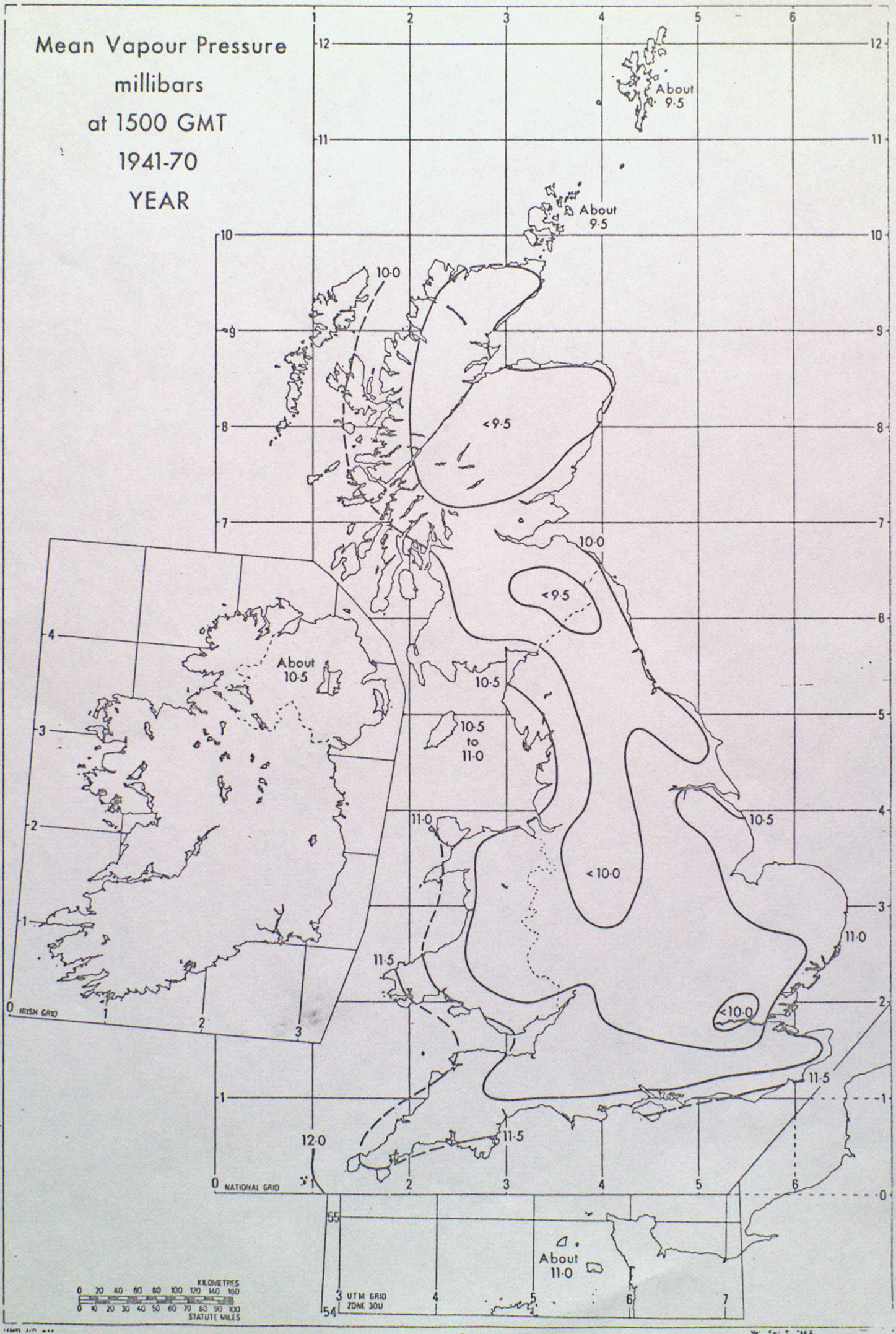
Meteorological Office, Met O 3
London Road
Bracknell
Berkshire
RG12 2SZ

Copies of Agricultural Memoranda are available from:

Meteorological Office, Met O 8a
London Road
Bracknell
Berkshire
RG12 2SZ

(Taken from Climatological Memorandum 72)




Mean Vapour Pressure
millibars
at 1500 GMT
1941-70
YEAR
About 9·5
About 9·5
10·0
<9·5
10·0
<9·5
About 10·5
10·5
10·5 to 11·0
11·0
10·5
<10·0
11·5
<10·0
11·0
11·5
12·0
11·5
About 11·0
IRISH GRID
NATIONAL GRID
UTM GRID ZONE 30U
KILOMETRES
STATUTE MILES

Hourly mean speed in metres per second exceeded for 50% of the time during the years 1965 - 73. Valid for a height of 10 metres above open level terrain (gust ratio 1·60) and for altitudes between 0 and 70 metres above mean sea level.